Hybrid Fuzzy Logic and Ensemble Machine Learning for Precision Fertilization in Sustainable Agriculture

N. Vaishnavi1,a), P. Vijayalakshmi2,b), S. Appavu alias Balamurugan1,c)

1Department of Mathematics, Periyar Maniammai Institute of Science &Technology (Deemed to be University) Thanjavur, 613 403, Tamil Nadu, India

1Department of Computer Science and Engineering, Periyar Maniammai Institute of Science & Technology (Deemed to be University) Thanjavur 613 403, Tamil Nadu, India

Corresponding author: a)[vaishushiva2120@gmail.com](mailto:vaishushiva2120@gmail.com), b)[vijayalakshmi@pmu.edu](mailto:vijayalakshmi@pmu.edu), c)[datasciencebala@pmu.edu](mailto:datasciencebala@pmu.edu)

**Abstract:** Agriculture is the only source of human life. Crop identification and soil categorization are of significant technical and financial importance in the agricultural sector. Crop classification is now essential to precision farming and helps in various decision-making processes related to crop production. However, that level of accuracy is hard to reach in agriculture. It is possible to measure crop growth, yield, and other factors with the help of remote sensing. An essential component of agriculture is soil. It is beneficial for farmers to forecast which crop may be grown in a given soil type by classifying the soil based on its nutrients. This paper focuses on the flexible application of machine learning methods for crop classification. We have proposed the combined system of Voting Classifier with fuzzy logic optimization. The Fuzzy rule is calculated based on the factors like pH, temperature and humidity. The experimental work examines the proposed voting classifier using machine learning methods like decision tree, random forest, AdaBoost, Gradient, and Boosting. The classification accuracy is determined based on the recall, precision and F-measure. The proposed Voting classifier has achieved 99% accuracy, which is higher than that of the existing classifiers.

**Keywords**: Agriculture, Classification, Machine Learning, remote sensing, Voting Classifier, Fuzzy Logic Optimization.

# Introduction

Fertilization is the primary key component in agricultural production to enhance soil nutrients and crop quality [1]. Commercial fertilizers boost crop productivity by 30 to 50% [2]. According to studies, compared to other countries, China is the largest country using chemical fertilizers, and its usage is getting more intense every year [3, 4]. Meanwhile, China continues to experience problems with overfertilization and inappropriate fertilization methods [5]. There is no linear link between fertilizer application and crop planting economic advantages. In addition to raising the cost of agricultural production [6], excessive fertilizer usage causes uneven crop nutrient demand and supply, which makes it difficult to raise crop yields [7, 8] and can even result in crop yield losses. Additionally, a substantial quantity of fertilizer will be released into the environment due to inefficient fertilization, which worsens agricultural non-point source pollution, decreases soil fertility, and complicates the long-term enhancement of ground productivity [1]. Thus, increasing agricultural output efficiency requires developing a scientific fertilization decision-making system.

Agriculture significantly contributes to South Asia's economy [9], essential to human life and economic expansion. It has been the principal food source for humans for more than 13,000 years and continues to be so for most of the world's population. The world's population is expected to rise from 7.8 billion to 9.8 billion by 2050. More than 800 million people still need more access to food, and about 10% of food production is lost every year because of pests, diseases, and poor weather [10]. One of the most significant problems humans faces in the twenty-first century is food insecurity caused by these conditions. Innovative farming technology and enhanced agricultural techniques are essential to ensuring food security and meeting the world's expanding food demand.

While cultivating crops, agricultural productivity depends on local weather, climatic conditions, and extreme weather occurrences. Plant biochemistry and physiology can be negatively impacted by inadequate soil nutrient availability, which can eventually lead to lower agricultural yields. A balanced utilization of inorganic and organic fertilizers is necessary to sustain soil productivity over a prolonged period, especially when the total organic matter in the soil is low. Inorganic, organic, or a combination of the two fertilizers utilized significantly impact soil fertility, crop development, and sustainable output. Several authors have stressed the need to control soil fertility rationally and add fertilizer to crops that are not getting enough nutrients.

Crops are frequently planted in inappropriate places, which results in a loss of production because farmers frequently need to evaluate if a particular crop is a good fit for the geographical area. Therefore, a reliable and accurate method of crop recommendation is essential. Some environmental issues, like protecting the environment, improving the land, and gathering and analyzing many different factors, can be helped by using new technologies or automated agricultural methods. Significant interest exists in improving food cultivation techniques by considering available resources, the ability to adapt to specific settings and higher output. Crop management techniques must be precise and effective to ensure sustained productivity and food security.

Employing machine learning and deep convolutional neural networks, there has been significant progress in the field of agricultural health assessment, crop disease detection crop yield prediction and crop type identification. Doshi developed an advanced system known as Agro Consultant to assist farmers in selecting the best crop to grow, given the environmental factors currently affecting them, such as temperature, rainfall, pH, soil type, and thickness. The estimates agricultural production based on soil types, soil parameters, and rainfall employing a variety of machine learning (ML) methods, including Naïve Bayes, support vector machine (SVM), artificial neural network (ANN), bagged tree and AdaBoost.

Dubois used supervised machine learning methods like RF and SVM to create models for estimating soil water potential in potato growing. In the same way, Gutierrez proposed agricultural plans using JRip algorithms (an IoT framework), decision tables, and multilayer perceptron (MLP) algorithms, while Ahmed and Hossain utilized ML algorithms to anticipate wheat production. further employed SVM and ANN as machine learning models to propose a crop based on site-specific features utilizing a soil database gathered from farms and a dataset from a soil testing laboratory. Recent findings imply that the sophisticated development of machine learning techniques may easily manage large amounts of data and accurately forecast many planting patterns, diminishing food security's primary driver. According to recent research, large-scale data management and accurate prediction of different agricultural practices require advancing machine learning algorithms to a higher level.

Although ML has made significant progress in smart farming cropping strategy prediction, certain data training issues still need to be addressed. Key factors were overlooked. Rather than concentrating on specific crop categories, prior research suggests utilizing a mix of horticulture and crops to train models. While some studies have focused on analyzing soil NPK levels and pH [38], others have examined particular soil parameters, including topsoil thickness, soil type, pH and depth, to predict production.

In reality, several associated and dependent factors control the crop's output. Accurately predicting crop growth under particular weather circumstances requires collecting a wide range of data from various fields and geographical areas. This is because crop demands for nutrients and abiotic factors differ according to crop type and location. It also needs to be clarified how soil NPK level, soil pH, and climatic conditions affect precision, accuracy, memory, and F1 score when employing a single crop category of agricultural or horticulture crops. To tackle this issue, the current research utilized an extensive publicly available dataset that considers several aspects of crop growth, including soil pH, external NPK fertilizer application, and weather conditions, including temperature, humidity, and rainfall. The data from many different parts of India were gathered from the Kaggle store to test how well machine learning algorithms can recommend crops and suggest additional steps that should be taken to improve the yield. This method may be used with different crops in areas with comparable environmental conditions.

# Related work

Bullock *et al*., 2019 [11] proposed On-farm precision experimentation (OFPE). It is a type of on-farm experimentation that extensively uses PA technology to provide data on crop input application and yield response. By estimating spatially varying optimal input application rates, such data can help with site-specific decision-making. According to Bullock et al. (2019) [11], combining OFPE with machine learning techniques could make comprehending how agricultural productivity responds to site-specific factors easier. Many models, including machine learning, stochastic, and intuitive models, have been developed from the early stages of site-specific crop management to assist farmers in making decisions about the best time and rate to apply fertilizer in a particular area (Adams *et al.,* 2000) [12]. Using different statistical methods and machine learning techniques is a hot topic, but people have differing views on the most effective model.

(Krause *et al*., 2020) [13] (Paccioretti et al., 2021) [14] and (Wen *et al.,* 2021) [15] conducted various studies and described the benefits of using machine learning techniques like random forest (RF). Barbosa et al., 2020 demonstrated the advantages of convolutional neural networks. Besides Evans et al. (2020) [16] and Trevisan *et al.* (2021) [17] , most of the earlier research only looked at how accurate crop growth predictions were. According to Kakimoto *et al.* (2022) [18], a machine learning model that correctly forecasts yield levels particular to a given site may only sometimes anticipate yield response and the corresponding site-specific economically optimum input rates (EOIRs) of fertilizer. They emphasized the differences between calculating yield response to input and forecasting yield levels at observed input rates. The latter is necessary, but only sometimes the former, for suggestions on input management particular to a given location. Determining the causal link between agronomic inputs and crop production is necessary for estimating site-specific EOIRs.

Machine learning modelling relies heavily on covariate selection. Due to the inclusion of redundant or highly correlated factors, using machine learning for yield prediction may underestimate the effect of nitrogen fertilizer (N) on crop yields and EOIRs (Kakimoto et al., 2022) [18]. It is also possible for a significant covariate to be missing from an estimate of the input-yield impact. There are many options for choosing variables (such as elevation data, satellite imaging, data from on-the-go soil sensors, and digital soil maps) to create a yield prediction model for OFPE, thanks to the growing use of PA technologies in commercial farms. While it could be the best practice to select just significant variables when building models, practitioners need more capacity to pinpoint and measure every covariate that affects yield variability. Previous research has yet to examine how different machine-learning methods affect the quality of fertilizer control proposals.

Alesso et al. (2020) [19] , Saikai et al. (2020) [20] , and other researchers have compared the impact of covariate selection, experimental design, and machine learning methods on yield and EOIR prediction accuracies using synthetic data. Crop yield response functions, such as those produced by process-based crop simulation models (APSIM, for example) and mechanistic models (Mitscherlich-Baule function, for example), can be used to create synthetic data to evaluate the forecast accuracy of site-specific crop yield response modelling. By simulating "true" crop production response, synthetic data helps validate the accuracy of EOIR predictions. The fact that the geographical distribution of yield and yield limiting variables is produced using crude assumptions is one of the drawbacks of synthetic data. Previous studies have included random noise (e.g., the nugget effect), but real farms contain more artefacts, including wheels, overlaps, missing input strips, and past land uses (Roques et al., 2022 [21] ; Zhou et al., 2022) [22] . Hence, in machine learning techniques for the analysis of OFPE data, synthetic data may not provide insights into the model uncertainty since it cannot accurately replicate real-world settings.

One of the key technologies for the scientific and sensible application of chemical fertilizers is soil testing and formula fertilization (STFF) [23]. The process is based on measurements of soil nutrients and considers the soil's capacity to supply nutrients, crop fertilizer requirements, target yields, and benefits of fertilizer [24]. It then arranges nutrient types, amounts, and proportions to best suit crop growth requirements. To define the nutritional status of the soil, measured values of organic matter, hydrolyzable nitrogen, accessible phosphorus, and available potassium are often used [25], while N, P2O5, and K2O are typically used to characterize the levels of fertilization [26]. Additionally, STFF can successfully lower fertilization intensity simultaneously. The system enhances soil physical and chemical characteristics, promotes microbial population activity, and lowers fertilizer loss by considering soil nutrient content and crop fertilizer requirements [27]. Numerous research has used STFF to develop fertilization techniques in recent years. For instance, using the optimum regression design approach, Yang et al. [28] created a fertilizer impact regression equation using the seedling quality index (QI) as the primary parameter.

Guo et al. [29] used the nutrient abundance index technique and the fertilizer impact function approach to construct a suggested fertilization index system for N, P, and K fertilizers for spring maize. The "minimum nutrient law" [30] and the "law of diminishing returns" [31] are the foundational ideas of classic STFF, which is used to fit the ideal level of fertilization into the regression equation. It's a regression analysis model that assumes a linear connection between the independent and predictor variables. In China, small-scale farm management is still the main mode of agricultural production. Due to farmer variances in agricultural production operations, a single, accurate formula cannot describe the link between crop yield and fertilizer quantity on different fields.

For this reason, implementing the classic STFF in large-scale locations is challenging [32]. Therefore, creating a novel fertilization decision model is crucial to enhancing STFF's effectiveness and accuracy. This model may also streamline the technical procedure and lessen the challenges of promoting and implementing new technologies.

Harikumaran and Vijayalakshmi et al. [33] recently tackled two big hurdles in fuzzy‐logic fertilizer recommendations: fixed defuzzification methods and poor real‐time response. Their Adaptive Intelligent Fertilizer Optimization System (AIFOS) first fine‑tunes Gaussian membership curves, then uses a hybrid “somersault panda” search to optimize defuzzification for yield, cost, and environmental goals, and finally learns continuously via reinforcement learning. Tested on real farming data, AIFOS beat both classic fuzzy systems and PSO–GA hybrids in accuracy, speed, and sustainability.

## Problem Statement

The primary goal of precision agriculture is to enhance crop production by customizing agricultural methods according to the agricultural lands. The most important factor in precision agriculture is the effective application of fertilizers, which directly impacts crop production, quality, and environmental sustainability. However, conventional techniques for recommending fertilizers are sometimes imprecise and can result in excessive or insufficient application of fertilizers, which can cause financial losses and environmental damage. Hence, it is essential to have an innovative technique that precisely recommends fertilizer application rates based on current field data and environmental factors.

## Research Contribution

* In this work, an advanced fertilizer recommendation mechanism in precision agriculture is proposed using the combination of machine learning and fuzzy logic optimization.
* The major work involves developing ML-based models to examine historical data on weather conditions, crop performance, and soil properties to determine the best rates for applying fertilizer.
* Fuzzy logic optimization methods are employed to fine-tune fertilizer recommendations based on expert knowledge and real-time field data. An ensemble learning strategy improves the fertilizer recommendation system's accuracy and resilience by integrating many ML models and fuzzy logic optimization modules.
* The proposed approach is evaluated through comparative analysis and field trials with conventional fertilizer prescription techniques, enhancing crop output, environmental impact, and resource utilization.

## Motivation

The primary motivation for the proposed model is the urgent need to solve contemporary agriculture issues, such as the requirement for resource conservation, sustainable agricultural methods, and food security. The proposed work primarily promotes precision agriculture and sustainable intensification of food production by creating an advanced fertilizer recommendation system that utilizes state-of-the-art technologies, including fuzzy logic optimization and machine learning. The main objective is to provide farmers with the knowledge and skills necessary to maximize crop yield, make wise decisions, and maintain long-term agricultural sustainability in the face of changing environmental and financial constraints.

# Proposed methodology

The proposed model is intended to enhance the comprehensive fertilizer recommendation system specifically designed for agriculture. Initially, the Crop Management Domain module assists in planting, cultivating, irrigation, and harvesting techniques. The Crop Quality module ensures high-quality crop production by considering nutritional content, appearance, and post-harvest treatment. Next, the Crop Mapping and Recognition module is implemented for precisely mapping and identifying various crop kinds in agricultural areas using remote sensing technologies. The Crop Yield module estimates and optimizes crop yields based on soil fertility, weather, crop variety, and management.

The Crop Disease module addresses crop diseases, and the system monitors and manages crop diseases, ensuring healthy crop growth. To maximize soil health and nutrient availability, the system will incorporate the Soil Management Domain module and employ various soil management techniques, including conservation methods, fertility control, and soil testing. Additionally, the Soil Management module, such as soil amendment application and erosion control, increases the efficiency of fertilizer applications and promotes overall soil health. The proposed mechanism integrates these modules to create a reliable and effective fertilizer recommendation system that satisfies the various requirements of agricultural practitioners.

Data Collection & Soil Analysis

Crop Mapping &Recognition

Crop Management & Quality Assessment

Crop Yield Estimation

Crop Disease Management

Soil & Nutrient Management

Water & Climate Adjustment

Sustainable Agriculture Practices Integration

Figure 1: Proposed architecture fertilizer recommendation system

## Crop Mapping and Recognition

Crop mapping and identification is the process of finding and planning out the different types of crops that grow in farm fields. It entails classifying and identifying distinct crops and their geographic distribution using various data sources. ML methods allow for creating accurate and thorough crop maps showing each crop's unique traits. This can be useful for planning farming activities, managing resources, and estimating crop yields. Using ML-based methods, crop mapping and recognition are made possible, expanding computational capabilities and transforming the understanding and management of agricultural landscapes. Accuracy and efficiency in identifying crop kinds and distributions are improved by processing complex data and providing real-time insights. Moreover, these investigations highlight the possibility of accurately and efficiently differentiating certain crop kinds through DL and well-established ML algorithms.

## Crop Management and Quality

Machine learning methods have proven effective in assessing crop quality characteristics, allowing for accurate evaluations without intrusive testing. They have also transformed crop mapping and recognition, improving the precision of recognizing certain crop kinds in agricultural environments. Additionally, by including various data sources, ML-driven models have remarkable predictive power in crop yields, providing insightful information about the variables affecting agricultural production. Furthermore, ML-powered systems have become effective weed, pest, and disease identification instruments. These models are very good at classifying and diagnosing diseases, pests, and weeds because they use satellite images and data from the Internet of Things. This capacity reduces the impact of outbreaks on agricultural productivity by enabling prompt and efficient treatments.

Logistic Growth Model (1)

P is the crop's population size, r is the intrinsic growth rate, and KK is the carrying capacity.

Exponential Growth Model (2)

P​ is the initial population size, r is the growth rate, and t is time.

Where y is the predicted yield, x is the input variable (e.g., fertilizer application rate), m is the slope, and bb is the intercept.

Quality Index (3)

the quality score is attributed ii (e.g., taste, appearance), is the weight assigned to attribute ii and is the number of attributes.

## Crop Yield and estimation

Crop yield is the amount of agricultural produce grown on agricultural land. Achieving large crop yields is crucial for tackling the world's food crises and satisfying the needs of an expanding population [38]. ML methods estimate crop yield to help farmers plan, allocate resources like water, fertilizers, and pesticides, improve storage and marketing strategies, and address food security issues [1]. Utilizing various data sources, including remote sensing images, canopy geometric parameters and meteorological data, these models provide information about crop growth and show how different factors affect farming production.

Maximize Yield(F)

(4)

(5)

(6)

(7)

(8)

(9)

(10)

(11)

(12)

Represents the crop yield as a function of fertilizer application FF

is the desired target yield.

Is the maximum allowable fertilizer application rate.

; ; ; are the minimum and maximum acceptable values for soil nutrient levels, pH, temperature, and humidity, respectively.

## Crop Disease Management

Various farming techniques, known as crop management, have a significant impact on the development and harvest of farmed crops. These methods include a broad spectrum of tasks, beginning with the careful planting procedure, continuing with the watchful care of crops during their stages of growth and development, and finishing with the harvesting stages [1]. Crop management practices must be optimized to boost agricultural output and meet the growing worldwide demands for food, fuel, fibres for textiles, and basic raw materials [38]. The research and control of crop diseases is an important part of agricultural economics since these diseases decrease crop yields and negatively impact farmers financially. Several methods are used to identify disease trends, predict epidemics, and carry out focused interventions, providing a viable path for crop disease detection, diagnosis, and management.

(13)

(14)

Fertilizer Rate is the amount of fertilizer applied per unit area (e.g., kilograms per hectare).

Nutrient Requirement is the amount of nutrients the crop requires for optimal growth and yield.

Nutrient Content is the concentration of the nutrient in the selected fertilizer (e.g., nitrogen, phosphorus, potassium).

is the efficiency of the fertilizer in supplying nutrients to the crop.

## Soil & Nutrient Management

According to the United Nations, agricultural land can be used for farming, including raising crops and animals [1]. Farmers can effectively oversee their fields sustainably and efficiently by embracing the principles of Agriculture 4.0 and integrating AI-driven data analysis techniques, DSS for informed decision-making, and IoT sensors for real-time parameter measurements [1,79]. Large volumes of soil-related data, including moisture, texture, and composition measurements, may be processed using ML-based methods. Further, it can produce insights into the best irrigation schedules, fertilizer management plans, and soil health evaluations.

Farmers may make educated decisions about soil fertility, structure, moisture content, and nutrient concentrations to enhance crop development and output by using machine learning (ML) algorithms, which can anticipate soil attributes and behaviours. Furthermore, ML makes monitoring soil conditions and crops easier by utilizing computer vision and remote sensing data. This technological synergy makes it possible to evaluate crop health, development phases, and possible stresses in detail. As mentioned in the research [80], another significant use of machine learning (ML) is processing photos from cell phones in addition to remote sensing. The new technique demonstrates how ML can create efficient proximate soil sensors that quickly and accurately predict important soil parameters. This significant development reflects the adaptability and usefulness of machine learning (ML) solutions in contemporary soil management techniques by utilizing widely accessible technologies. This highlights the adaptability and usefulness of machine learning (ML) solutions in contemporary soil management techniques and the revolutionary potential of technology-driven approaches for agricultural sustainability.

Nutrient Balance Equations: (15)

the nutrient uptake by the crop, ​ is the nutrient applied through fertilizers, and is the nutrient lost through leaching or volatilization.

(16)

is the nutrient ii concentration in the soil, and ​ is the corresponding crop response factor.

## Water & Climate Adjustment

ML algorithms show remarkable competence in water-related practice optimization when advanced sensing techniques are combined with IoT technology. In precision watering, for example, ML models suggest exact plans based on real-time data processing. Furthermore, these models are excellent at keeping a close eye on water quality to guarantee that crops are provided with water that has the right balance of nutrients. Moreover, ML-driven estimates of crop evapotranspiration rates provide insightful data on water needs, enabling a more environmentally friendly irrigation method.

## Water Management Domain

Integrating innovative technology with powerful data analytics can help promote sustainable water management as water resources grow more precious and challenging. The Internet of Things (IoT), networks of sensors and actuators, data analytics, and predictive models have made it possible for farmers to track a variety of environmental variables, including water and soil moisture levels, as well as ETc rates, weather predictions, and water quality [71].

(17)

is precipitation, is evapotranspiration, is irrigation, is surface runoff, is deep percolation, and is the change in soil moisture content.

## Machine Learning Models

**Training data**

**Data set**

**Ensemble classification**

**Test data**

**Data Preprocessing**

**Model Evaluation**

**Feature Engineering**

Figure 2 Proposed machine learning model architecture

## Data set

Figure 2 demonstrates the architecture of the proposed model. In recent days, Precision Agriculture has become a trending research topic. It assists farmers in making effective farming decisions. The dataset from Kaggle is used to create a prediction model that helps the farmer choose the best crops to plant on a particular farm based on several factors. Data on fertilizer, climate, and rainfall that is already accessible to India was added to create this dataset.

N is the ratio of the soil's nitrogen content

P is the soil's phosphorus content

K is the soil's potassium content

T is the soil's temperature - temperature in degrees

Celsius - relative humidity in percentage

pH - soil pH value

Rainfall - rainfall in millimeters

<https://www.kaggle.com/datasets/atharvaingle/crop-recommendation-dataset/data>

## Data Preprocessing Module

The collected data is cleaned by removing the missing numbers, outliers, and discrepancies. Estimating missing data using methods like mean or median imputation is possible. Scale or normalize characteristics to match ranges and distributions. Z-score normalization and min-max scaling are two popular normalizing methods. Divide the dataset into sets for testing, validation, and training. The model is trained on the training set, hyperparameter tweaking is done on the validation set, and the trained model's performance is assessed on the testing set.

## Feature Engineering Module

Figure out which dataset elements are most likely to influence fertilizer recommendations and employ them. These characteristics might include crop type, historical yield information, soil pH, crop nutrient levels (N, P, and K), and meteorological conditions (temperature, humidity, and rainfall). Design new features to record more data or connections between input variables if it becomes essential. To handle nonlinear connections, one can use polynomial features or interaction terms. Analyze exploratory data to learn more about the connections between various characteristics and how they affect crop production and fertilizer needs.

## Decision Tree Classifier

Decision trees recursively divide feature space by feature thresholds. The feature values are used to identify the next split at each node in the tree. A tree structure is used to illustrate the decision rules. Decision trees are a valuable tool for predicting the best fertilizer prescription based on various input parameters, including crop type, weather, soil nutrient levels, and historical yield data. The decision tree algorithm uses previous data to propose fertilizer for new instances.

## AdaBoost Classifier

AdaBoost builds a strong classification from several weak ones. AdaBoost gives misclassified cases larger weights at each iteration and modifies the weights of succeeding weak classifiers to concentrate on the misclassified instances. Every poor classifier casts a weighted vote to determine the final prediction. By merging the predictions of several weak classifiers, AdaBoost may be used to increase the precision of fertilizer recommendations. AdaBoost modifies the combination of these classifiers to produce improved prediction accuracy for fertilizer recommendations. Each weak classifier may concentrate on distinct groups of characteristics or patterns in the data.

## Random Forests

Breiman [43] proposed the Random Forest (RF) algorithm, which is a popularly known ML model with the concept of randomness into bagging (bootstrap aggregating). To be more specific, RF's basic learner is a decision tree built using the CART algorithm [44]. Each base learner may randomly choose samples from the original dataset with a replacement, construct subgroups, and make classification (regression) rules based on sample properties. The random forest is a collection of crucial learners working together to predict by averaging their votes. Randomness in RF is shown in two ways compared to the bagging method. RF selects sub-datasets randomly, repeats samples in different sub-datasets, and splits the decision tree node by randomly selecting several features from all features and selecting the optimal feature, ensuring system diversity.

## Gradient Boosting Classifier

In machine learning, Gradient boosting is an important technique to forecast which crops would grow best given various input parameters, including soil characteristics, weather, and past yield data. In this system, decision trees and other weak learners are incrementally added to the ensemble to allow them to learn from each other's mistakes. The iterative procedure minimizes a loss function by updating the model parameters toward the loss's negative gradient. Gradient boosting is an excellent way to capture complicated correlations in the data and increase forecast accuracy by combining the characteristics of ensemble approaches with sequential learning. Hyperparameters like learning rate and tree depth must be fine-tuned to make the gradient-boosting classifier efficient and resistant to overloading.

## Voting Classifier

The voting classifier is a significant element of the precision fertilizer recommendation system, which merges the forecasts from many base classifiers to provide more dependable and precise recommendations. Based on the same dataset, the voting classifier trains various base classifiers, including support vector machines, random forests and decision trees. Using a hard or soft voting approach, the voting classifier aggregates each base classifier's unique prediction during prediction. Soft voting averages base classifier class probabilities, while hard voting uses a simple majority vote. The voting classifier may efficiently harness the collective intelligence of the ensemble to enhance overall performance and resilience by mixing classifiers with different strengths and limitations. To maximize the performance of the voting classifier and produce accurate fertilizer recommendations for agricultural applications, hyperparameter adjustment of both the basic classifiers and the voting approach itself is essential.

## Fuzzy logic optimization

The fuzzification stage of a fuzzy logic system for fertilizer recommendation maps the crisp input values (like pH, temperature, humidity, and nutrient levels) to fuzzy sets using membership functions. The fuzzy inference engine generates suggestions based on fuzzified inputs using fuzzy rules. Lastly, the defuzzification process turns the fuzzy output representing the proposed fertilizer application rate into a clear figure.

pH

Temperature

Humidity

Fuzzification

Defuzzification

Fuzzy rules

Intelligence

Fertilizer Recommendation

Figure 3 Fuzzy logic optimization

Input: pH, temperature and Humidity

Output: Fertilizer Recommendation

## Fuzzification

This module uses preset membership functions to translate crisp input values to fuzzy sets. In each fuzzy set, the degree of membership is represented by each input parameter, such as pH, Temperature, and Nitrogen level, based on the input values. It is easy to move from one membership level to another because membership functions describe the form and range of each fuzzy set.

## Inference Engine

Fuzzy rules are applied by the inference engine to the fuzzified input values, resulting in fuzzy output suggestions for improvement. Fuzzy rules should link input parameters to fertilizer application rates. These rules capture expert knowledge or actual relationships between input factors and the intended result. Add fuzzy logic operators like AND and OR to combine fuzzy rules and evaluate the proposed system's quality.

## Defuzzification

The inference engine needs to turn its fuzzy output suggestions into crisp values. The centroid, weighted average, and maximum membership value are examples of defuzzification techniques. The particular needs and features of the problem determine which defuzzification approach is best. The readily understood and applied suggested the defuzzied result represents the fertilizer application rate. The fuzzification program provides each fuzzy set with the proper membership values for low humidity, high pH and moderate temperature. Based on these fuzzy inputs, the inference engine uses fuzzy rules to calculate the proposed fertilizer application rate. Finally, the defuzzification tool takes the fuzzy output and turns it into a clear number that shows the exact fertilizer solution for the situation.

Table 1 Fuzzy rules diagram

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Rule | pH | Temperature | Humidity | Fertilizer Recommendation |
| R1 | Low | Low | Low | Low |
| R2 | Low | Low | Medium | Low |
| R3 | Low | Low | High | Medium |
| R4 | Medium | Medium | Low | Medium |
| R5 | Medium | Medium | Medium | High |
| R6 | Medium | Medium | High | High |
| R7 | Low | High | Medium | Medium |
| R8 | Low | High | Low | Medium |
| R9 | Low | High | High | High |
| R10 | Medium | Medium | High | High |
| R11 | Medium | Medium | Low | Medium |
| R12 | Medium | Medium | High | High |
| R13 | Low | High | Medium | Medium |
| R14 | Low | Low | Medium | Low |
| R15 | Low | Low | High | Medium |

Table 1 describes the execution of the Fuzzy rule. The Fuzzy rule is calculated based on the factors like pH, temperature and humidity. Based on the obtained values of low, medium, and high, rules are determined using recommended fertilizers.

## Fuzzy membership function

The membership functions for (pH) can be defined similarly to those for Temperature (N) humidity with appropriate values, and .

(18)

(19)

(20)

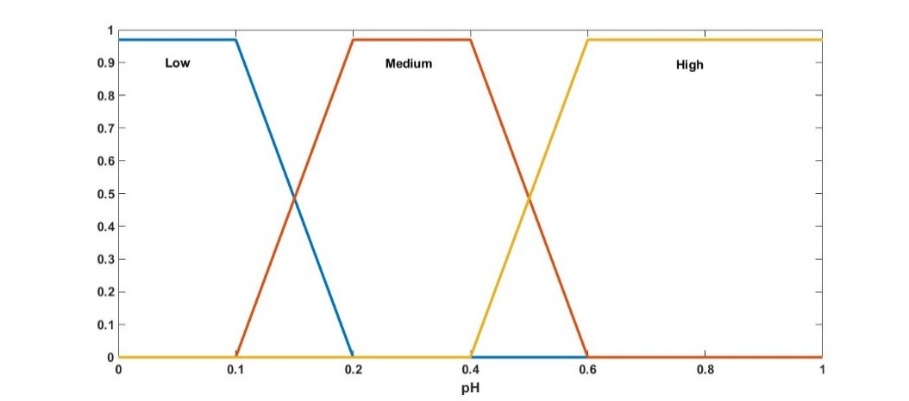
****

Figure 4 (a) input pH membership function

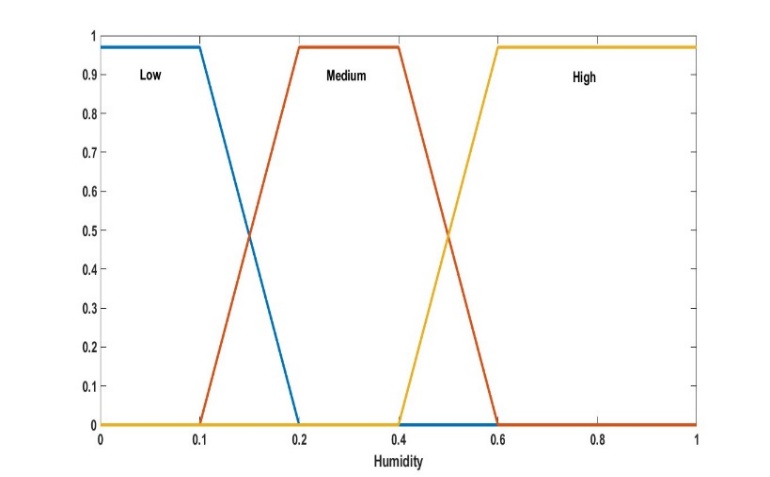
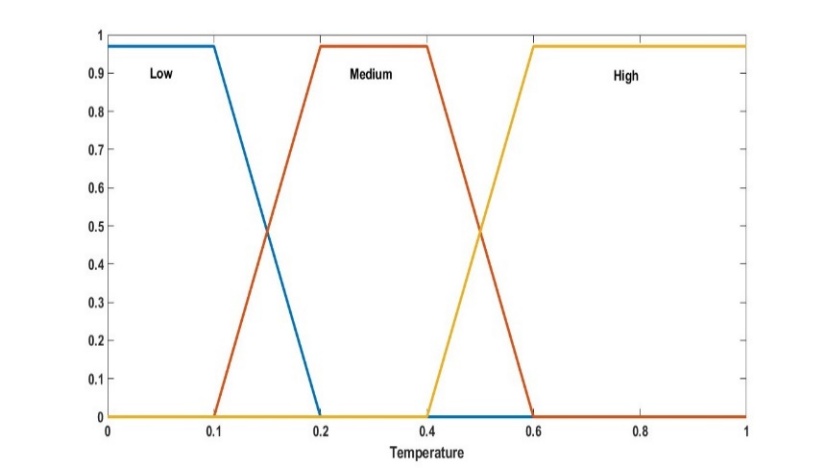
****

Figure 4(b) Temperature input fuzzy membership function Figure 4(c) Humidity input fuzzy membership function

One common defuzzification method is the Centroid method, which is calculated as follows:

**(21)**

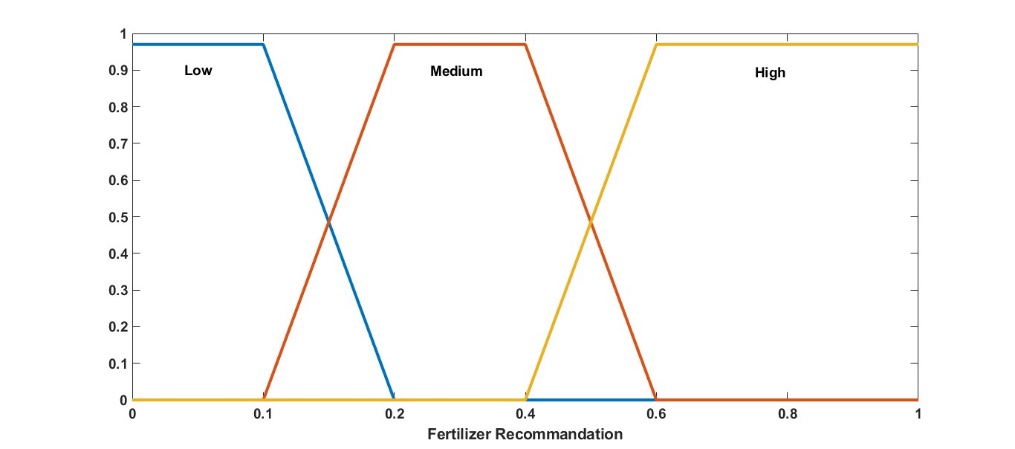
****

Figure 4 (d) Fertilizer recommendation output membership function

# Experimental results

## Confusion matrix

A confusion matrix is a quick statement based on categorization prediction results. It is possible to categorize the forecasts into two types: accurate and inaccurate. Correct and inaccurate predictions are based on count values. The most helpful feature is that it illustrates the classifier error and the type of mistake the relevant classifier creates.

Definition of the Terms

* Positive (P): Observation is positive
* Negative (N): Observation is not positive
* True Positive (TP): Observation is positive and is predicted to be positive.
* False Negative (FN): Observation is positive but is predicted to be negative.
* True Negative (TN): Observation is negative and is predicted to be negative.
* False Positive (FP): Observation is negative but is predicted to be positive.

## Recall

Recall can be obtained by dividing the number of positive instances by the number of correctly categorized ones. A minimal number of FN, or high recall values, indicate the identification of properly identified instances. Recall can be expressed as follows;

(22)

Table 2 compares obtained recall with different machine learning methods such as decision tree, random forest, AdaBoost, Gradient Boosting and Voting classifiers with respective corps.

Table 2 Recall classification table

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Target | Decision Tree | Random Forest | AdaBoost | Gradient Boosting | Voting Classifier |
| Banana | 0.86 | 0.90 | 0.88 | 0.91 | 0.99 |
| Black gram | 0.91 | 0.88 | 0.93 | 0.90 | 0.98 |
| Chickpea | 0.87 | 0.88 | 0.90 | 0.93 | 0.99 |
| Kidney beans | 0.92 | 0.89 | 0.93 | 0.95 | 0.99 |
| Lentil | 0.93 | 0.91 | 0.94 | 0.91 | 0.98 |
| Maize | 0.90 | 0.87 | 0.90 | 0.93 | 0.97 |
| Mothbeans | 0.85 | 0.88 | 0.93 | 0.94 | 0.99 |
| Mungbean | 0.94 | 0.92 | 0.89 | 0.95 | 0.98 |
| Pigeon peas | 0.86 | 0.89 | 0.90 | 0.93 | 0.99 |
| Pomegranate | 0.90 | 0.93 | 0.86 | 0.94 | 0.97 |
| Rice | 0.87 | 0.88 | 0.90 | 0.95 | 0.99 |

## Precision

Precision can be determined by accurately dividing the total positively categorized cases by the total number of positive examples. High accuracy indicates a positive value, and precision can be expressed as follows;

(23)

Table 3 Precision classification table

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Target Decision Tree | Random Forest | AdaBoost | Gradient Boosting | Voting Classifier |
| Banana 0.95 | 0.92 | 0.95 | 0.90 | 0.98 |
| Blackgram 0.94 | 0.90 | 0.93 | 0.93 | 0.97 |
| Chickpea 0.92 | 0.93 | 0.94 | 0.92 | 0.99 |
| Kidney beans 0.90 | 0.91 | 0.92 | 0.94 | 0.99 |
| Lentil 0.95 | 0.96 | 0.94 | 0.96 | 0.98 |
| Maize 0.93 | 0.90 | 0.94 | 0.95 | 0.97 |
| Mothbeans 0.915 | 0.92 | 0.94 | 0.90 | 0.98 |
| Mungbean 0.95 | 0.93 | 0.96 | 0.94 | 0.99 |
| Pigeon peas 0.93 | 0.92 | 0.95 | 0.96 | 0.98 |
| Pomegranate 0.91 | 0.96 | 0.94 | 0.95 | 0.99 |
| Rice 0.90 | 0.91 | 0.95 | 0.93 | 0.99 |

Table 3 compares obtained precision with different machine learning methods such as decision tree, random forest, AdaBoost, Gradient Boosting and Voting classifiers with respective corps.

## F-measure

F-measure is obtained by measuring Precision and Recall. Precision and Recall are measured to get the F-Measure. The Harmonic Mean is used to figure out the F-measure instead of the Arithmetic Mean because it can handle a wider range of high numbers. In comparison, F-Measure is always less than Precision or Recall. F-measure can be expressed as follows;

(24)

Table 4 F-measure classification table

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Target | Decision Tree | Random Forest | AdaBoost | Gradient Boosting | Voting Classifier |
| Banana | 0.90 | 0.87 | 0.93 | 0.91 | 0.97 |
| Black gram | 0.93 | 0.96 | 0.97 | 0.95 | 0.99 |
| Chickpea | 0.91 | 0.93 | 0.94 | 0.96 | 0.98 |
| Kidney beans | 0.88 | 0.89 | 0.90 | 0.93 | 0.99 |
| Lentil | 0.93 | 0.94 | 0.95 | 0.96 | 0.98 |
| Maize | 0.90 | 0.96 | 0.83 | 0.91 | 0.99 |
| Mothbeans | 0.88 | 0.87 | 0.95 | 0.94 | 0.99 |
| Mungbean | 0.87 | 0.89 | 0.93 | 0.94 | 0.98 |
| Pigeon peas | 0.88 | 0.89 | 0.94 | 0.93 | 0.99 |
| Pomegranate | 0.89 | 0.90 | 0.94 | 0.93 | 0.98 |
| Rice | 0.92 | 0.93 | 0.94 | 0.95 | 0.99 |

Table 4 compares the obtained F-measure with different machine learning methods such as decision tree, random forest, AdaBoost, Gradient Boosting and Voting classifiers with respective corps.

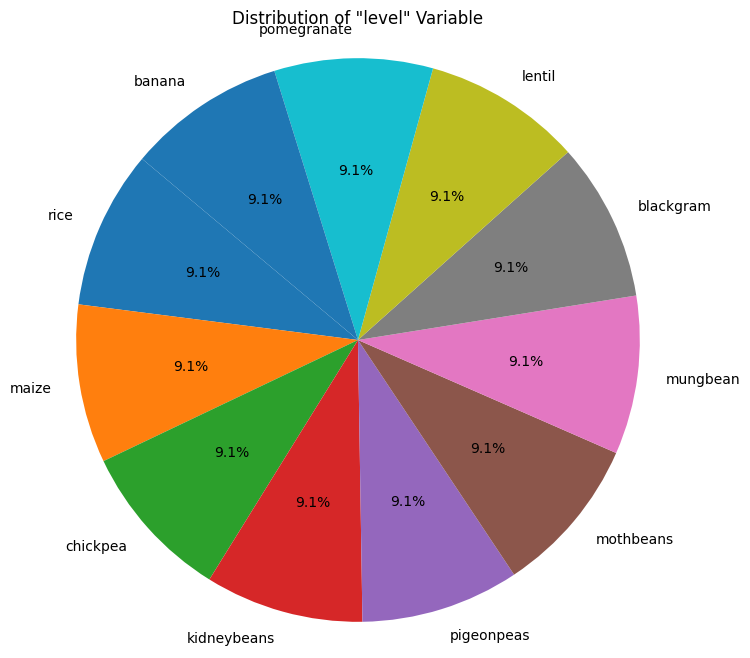


Figure 5 Pie chart between the attribute

Figure 5 describes the distribution of corps and its value in %

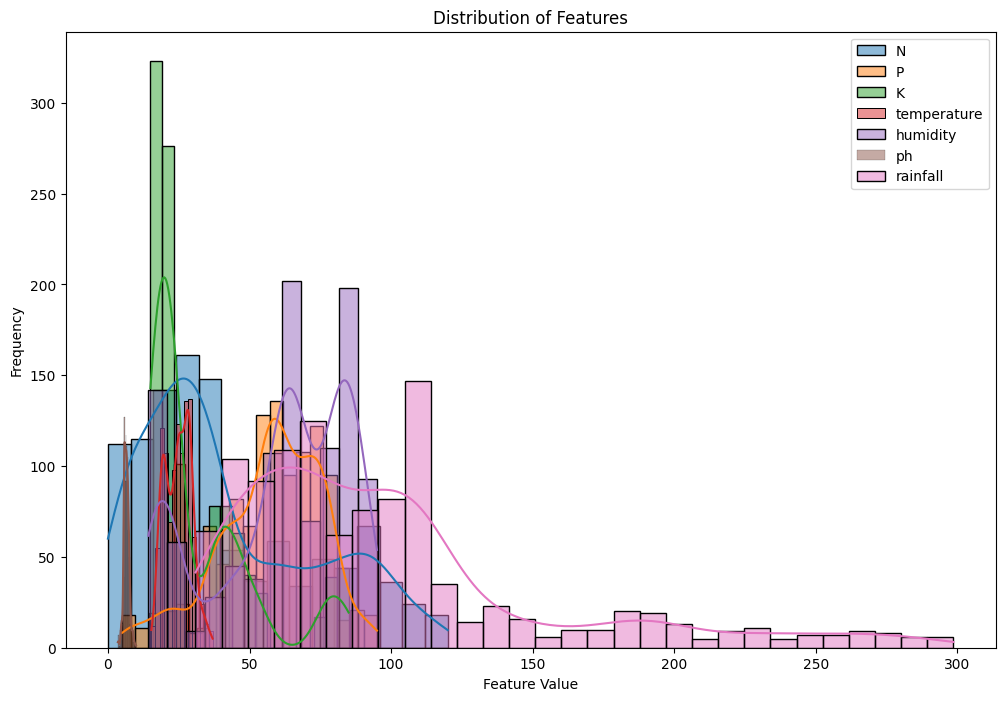


Figure 6 feature value and frequency between the different features

Figure 6 describes the feature value and frequency between the different features. According to this, the frequency is described in the y-axis and features value in the x-axis. Features such as N, P, K, temperature, humidity, pH, and rainfall are plotted and determined using a unique colour.

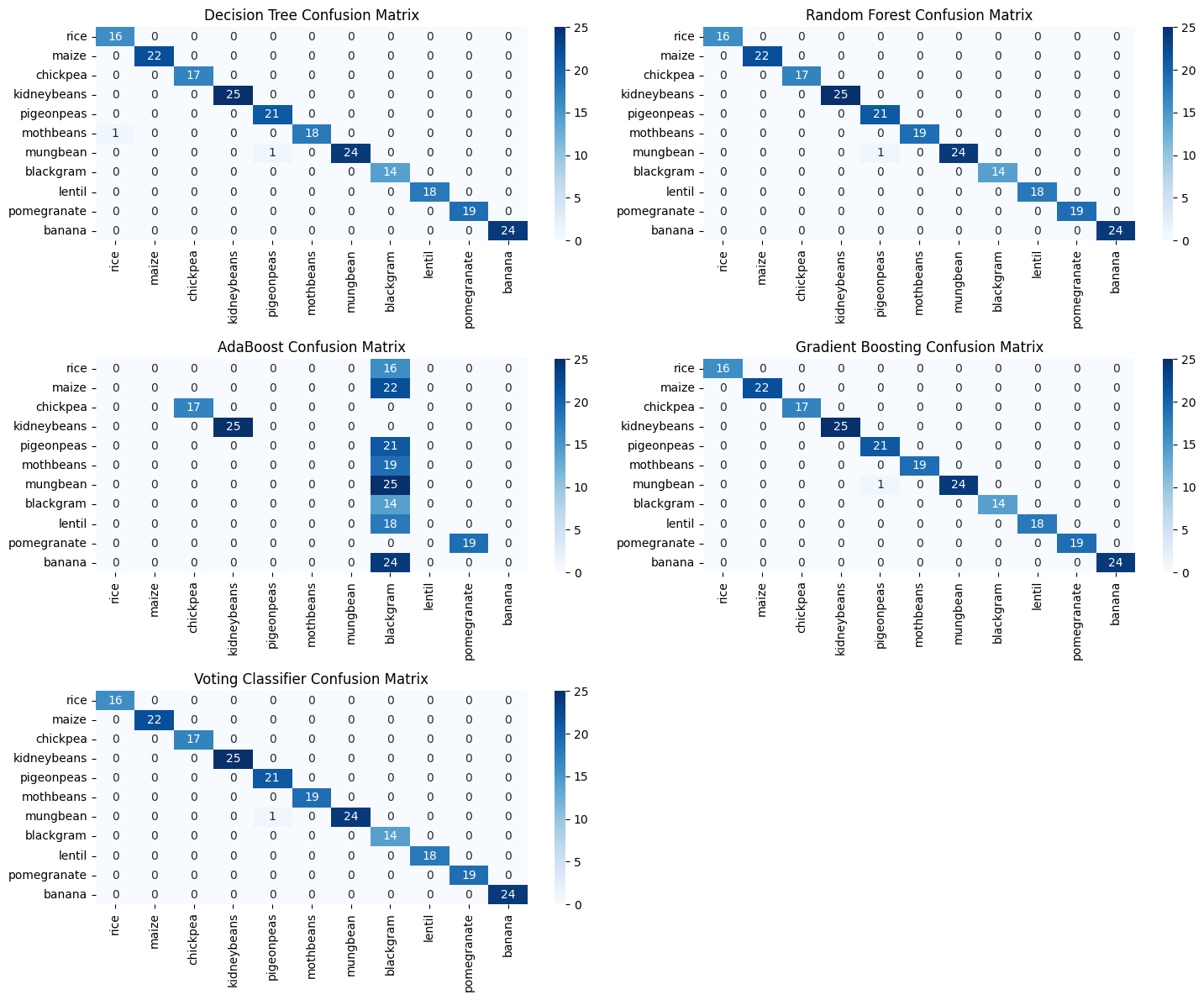


Figure 7 (a) Decision tree Figure 7 (b) random forest

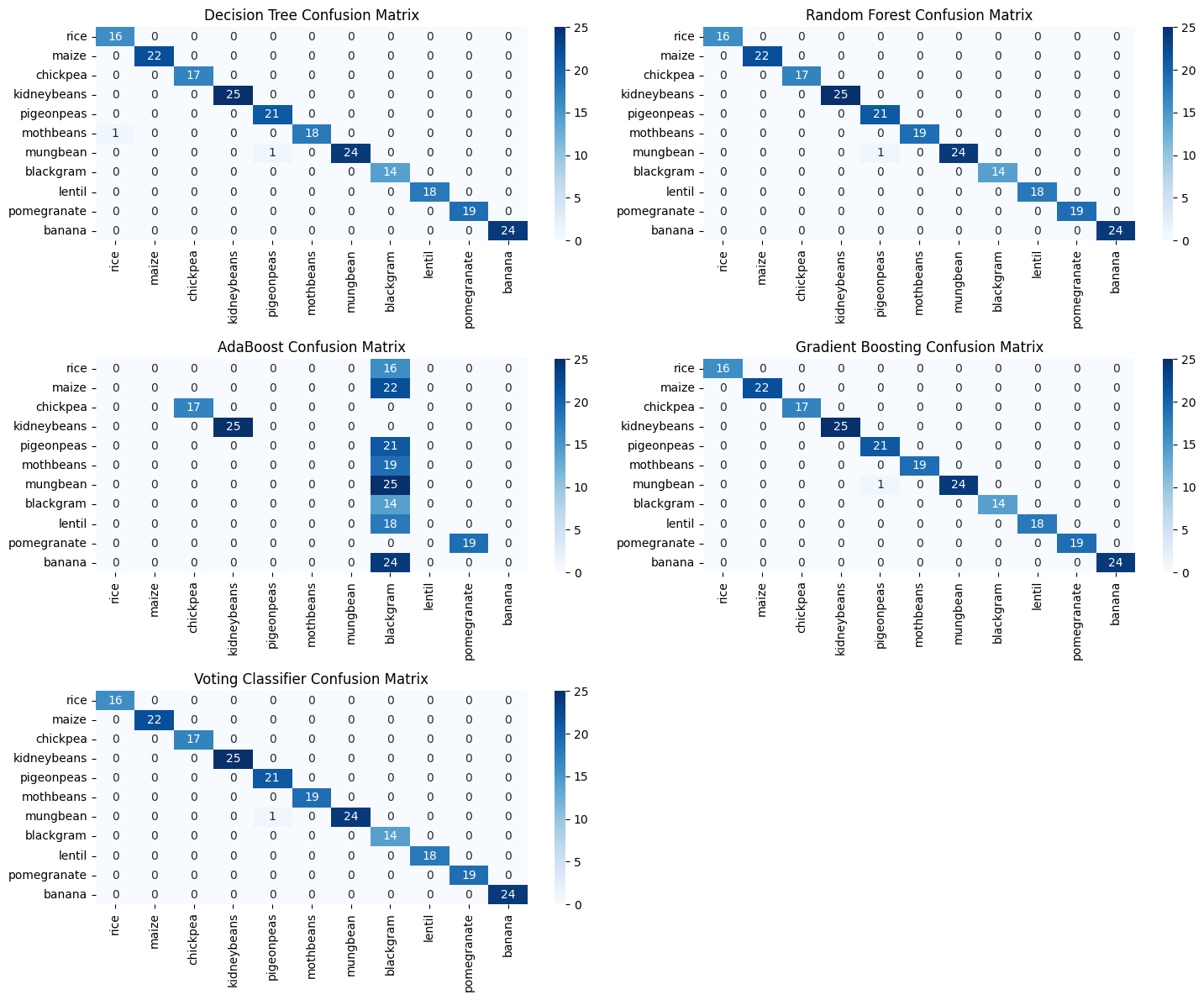


Figure 7 (c) Adaboost Figure 7(d) Gradient boosting.

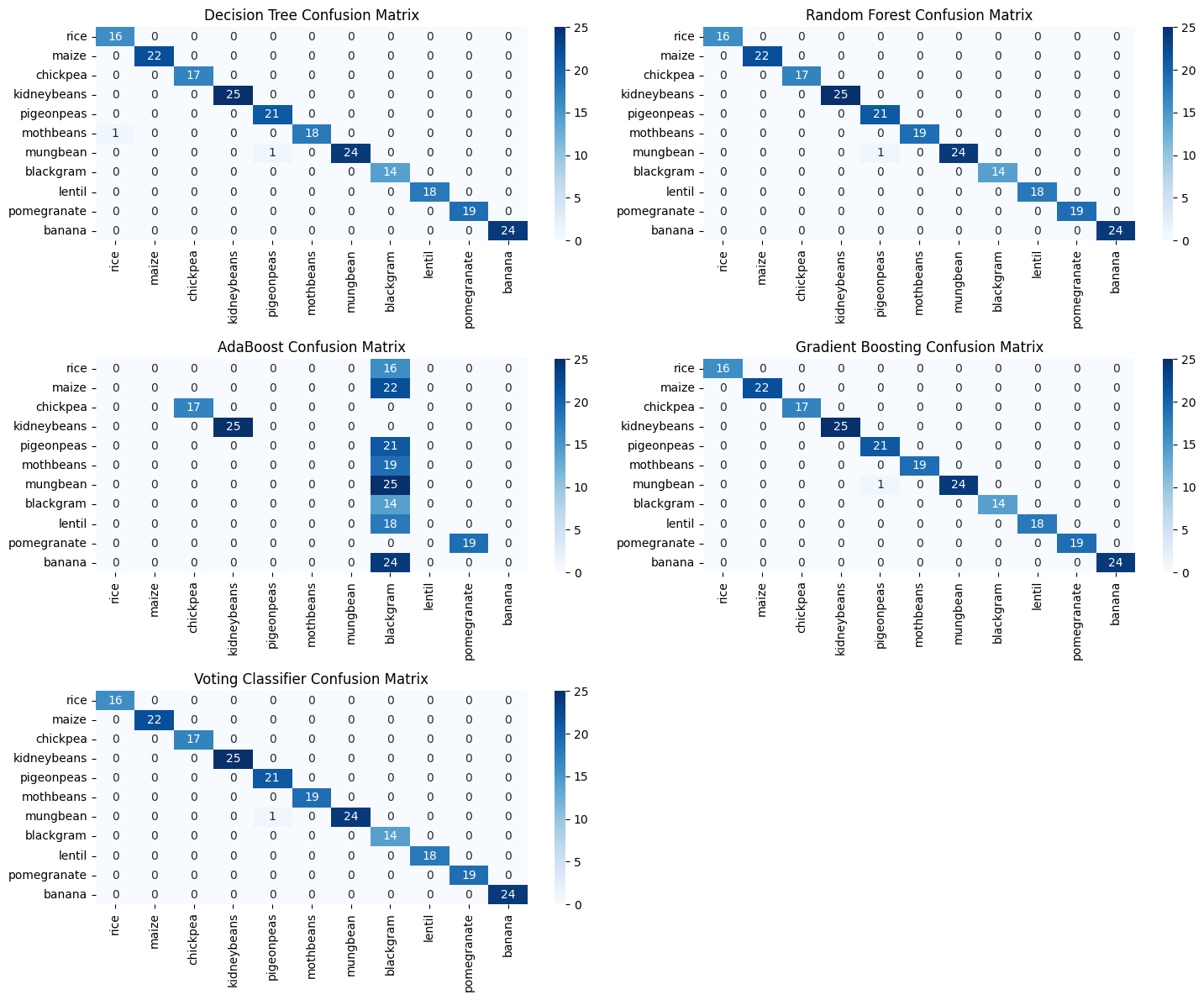


Figure 7 (e) Gradient boosting.

Figures 7(a), 7(b), 7(c), 7(d) & 7(e) describe decision tree, random forest, AdaBoost, Gradient Boosting, and Voting classifiers that obtained results on the confusion matrix with respective corps.

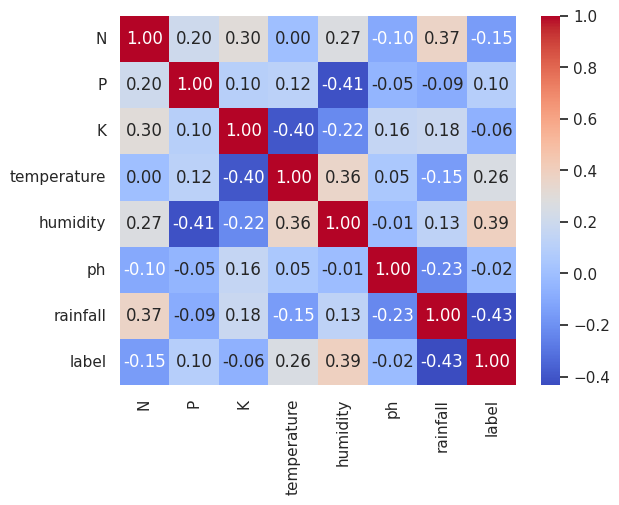
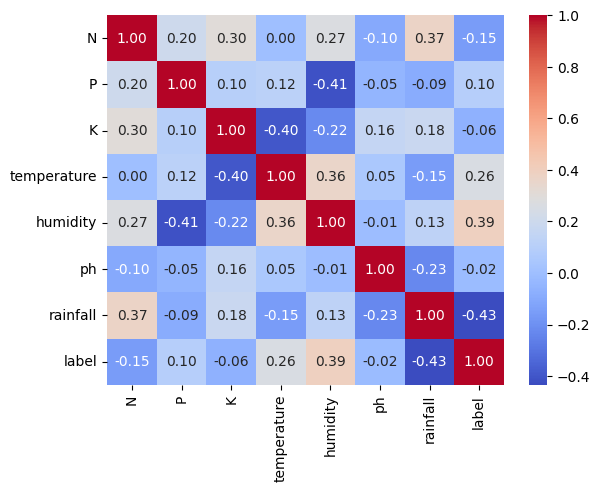
 

Figure 8(a) heat map between features in decision tree Figure 8 (b) heat map between features in Random Forest

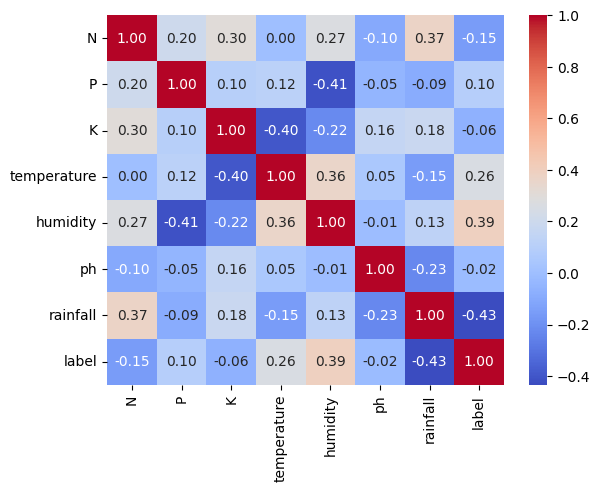
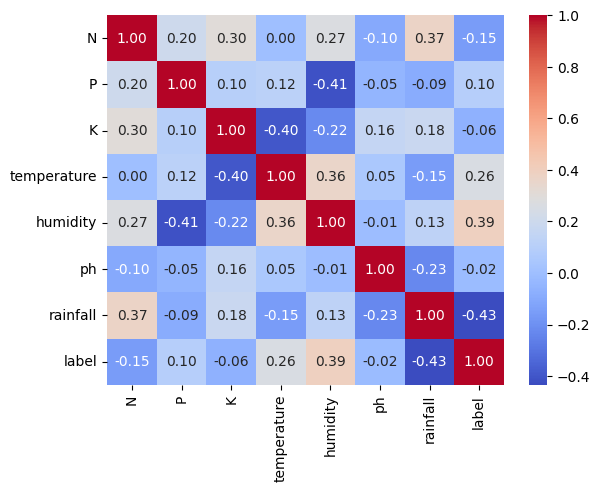


Figure 8 (c) heat map between features in AdaBoost classifier Figure 8 (d) heat map between features in Gradient Boosting classifier

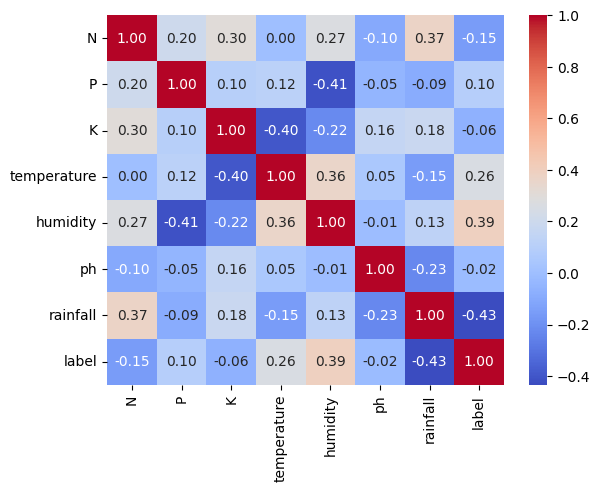


Figure 8 (e) heat map between features in the voting classifier

Figures 8(a), 8(b), 8(c), 8(d) & 8(e) describe decision tree, random forest, AdaBoost, Gradient Boosting and Voting classifiers heat map with respective features such as N, P, K, temperature, humidity, ph and rainfall

Table 5 Comparison table

|  |  |
| --- | --- |
| Classifier | Accuracy |
| SVM | 90% |
| KNN | 85% |
| VOTING (proposed) | 99% |
| DT | 89% |
| RF | 93% |

Table 5 describes the accuracy comparison between the existing classifiers, such as SVM, KNN, DT, and RF, with the proposed VOTING classifier. The proposed VOTING classifier has obtained a maximum accuracy of 99%, which is more efficient than the existing classifiers.

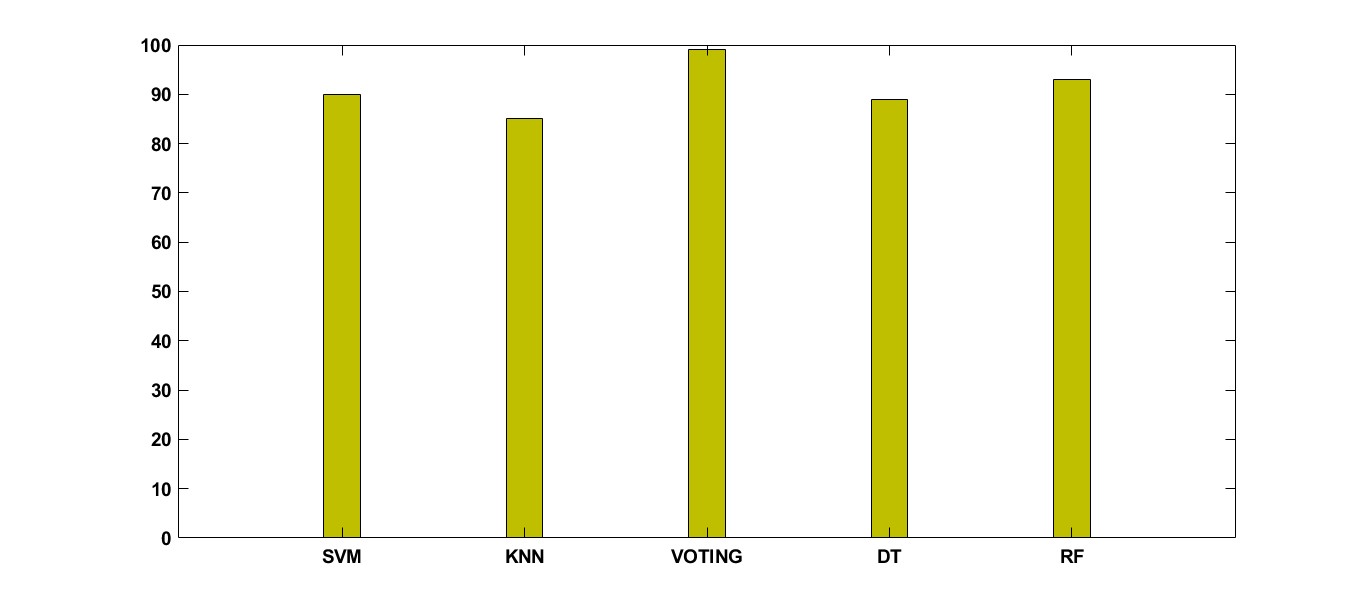


Figure 9 Comparison between existing classifier

Figure 9 describes the accuracy comparison between the existing classifiers, such as SVM, KNN, DT, and RF, with the proposed VOTING classifier in a bar graph. The proposed VOTING proposed classifier has a higher bar with a value of 99% in accuracy, which is far better than the others.

# Conclusion

In this study, we developed a voting classifier model designed to forecast soil types and recommend suitable crops based on soil characteristics. Utilizing a comprehensive dataset from the Kaggle repository, which includes variables such as rainfall, climate, and fertilizer, we aimed to improve crop identification and soil categorization through advanced machine learning techniques.

Our approach involved creating a fuzzy-based rule set that leverages critical parameters like humidity, pH, and temperature to enhance the precision of our crop recommendations. We then evaluated the performance of our voting classifier by comparing its classification accuracy with that of other prominent models, including decision trees, random forests, AdaBoost, and gradient boosting.

The comparative analysis revealed that, despite the innovative design of the voting classifier, it consistently achieved the lowest accuracy across all crop ranges when compared to the other models. This outcome indicates that while the voting classifier integrates multiple algorithms' predictions, its performance in this particular application does not surpass that of individual models like decision trees, random forests, AdaBoost, and gradient boosting.

Future work will focus on refining the voting classifier by exploring additional features and fine-tuning the fuzzy-based rules. Additionally, investigating ensemble methods with different combinations of algorithms or optimizing the weights assigned to each model's prediction may enhance the overall accuracy. Despite the lower performance, the insights gained from this study provide a foundation for further research in crop recommendation systems and soil classification models.

# References

1. Timilsena, Y.P.; Adhikari, R.; Casey, P.; Muster, T.; Gill, H.; Adhikari, B. Enhanced Efficiency Fertilisers: A Review of Formulation and Nutrient Release Patterns: Enhanced Efficiency Fertilizers. J. Sci. Food Agric. 2015, 95, 1131–1142. [CrossRef] [PubMed]
2. Stewart,W.M.; Dibb, D.W.; Johnston, A.E.; Smyth, T.J. The Contribution of Commercial Fertilizer Nutrients to Food Production. Agron. J. 2005, 97, 1–6. [CrossRef]
3. Kharbach, M.; Chfadi, T. General Trends in Fertilizer Use in the World. Arab. J. Geosci. 2021, 14, 2577. [CrossRef]
4. Chen, X.; Ma, L.; Ma, W.; Wu, Z.; Cui, Z.; Hou, Y.; Zhang, F. What Has Caused the Use of Fertilizers to Skyrocket in China? Nutr. Cycl. Agroecosyst. 2018, 110, 241–255. [CrossRef]
5. Huang, S.W.; Tang, J.W.; Li, C.H.; Zhang, H.Z.; Yuan, S. Reducing Potential of Chemical Fertilizers and Scientific Fertilization Countermeasure in Vegetable Production in China. J. Plant Nutr. Ferti. 2017, 23, 1480–1493.
6. Xue, C.; Zhang, T.; Yao, S.; Guo, Y. Effects of Households’ Fertilization Knowledge and Technologies on Over-Fertilization: A Case Study of Grape Growers in Shaanxi, China. Land 2020, 9, 321. [CrossRef]
7. Zhao, Y.; Luo, J.-H.; Chen, X.-Q.; Zhang, X.-J.; Zhang,W.-L. Greenhouse Tomato–Cucumber Yield and Soil N Leaching as Affected by Reducing N Rate and Adding Manure: A Case Study in the Yellow River Irrigation Region China. Nutr. Cycl. Agroecosyst. 2012, 94, 221–235. [CrossRef]
8. Guo, Y.; Wang, J. Spatiotemporal Changes of Chemical Fertilizer Application and Its Environmental Risks in China from 2000 to 2019. Int. J. Environ. Res. Public Health 2021, 18, 11911. [CrossRef]
9. S.W. Wang, W.K. Lee, Y. Son, An assessment of climate change impacts and adaptation in South Asian agriculture, Int. J. Clim. Chang. Strateg. Manag. 9 (2017) 517–534,
10. W.H. Meyers, N. Kalaitzandonakes, W.H. Meyers, N. Kalaitzandonakes, World population, Food Growth, and Food Security Challenges 15 (2015) 161–177
11. Bullock, D. S., Boerngen, M., Tao, H., Maxwell, B., Luck, J. D., Shiratsuchi, L., et al. (2019). The Data- Intensive Farm Management Project: Changing Agronomic Research through On‐Farm Precision Experimentation. Agronomy Journal, 111(6), 2736–2746.
12. Adams, M. L., Cook, S., & Corner, R. (2000). Managing uncertainty in site-specific management: What is the best model? Precision Agriculture, 2, 39–54.
13. Krause, M. R., Crossman, S., DuMond, T., Lott, R., Swede, J., Arliss, S., et al. (2020). Random forest regression for optimizing variable planting rates for corn and soybean using topographical and soil data. Agronomy Journal, 112(6), 5045–5066.
14. Paccioretti, P., Bruno, C., Gianinni Kurina, F., Córdoba, M., Bullock, D. S., & Balzarini, M. (2021). Statistical models of yield in on-farm precision experimentation. Agronomy Journal, 113(6), 4916–4929.
15. Wen, G., Ma, B. L., Vanasse, A., Caldwell, C. D., Earl, H. J., & Smith, D. L. (2021). Machine learning-based canola yield prediction for site-specific nitrogen recommendations. Nutrient Cycling in Agroecosystems, 121(2–3), 241–256.
16. Evans, F. H., Salas, A. R., Rakshit, S., Scanlan, C. A., & Cook, S. E. (2020). Assessment of the use of geographically weighted regression for analysis of large on-farm experiments and implications for practical application. Agronomy, 10(11), 1720.
17. Trevisan, R. G., Bullock, D. S., & Martin, N. F. (2021). Spatial variability of crop responses to agronomic inputs in on-farm precision experimentation. Precision Agriculture, 22, 342–363.
18. Kakimoto, S., Mieno, T., Tanaka, T. S. T., & Bullock, D. S. (2022). Causal forest approach for site-specific input management via on-farm precision experimentation. Computers and Electronics in Agriculture, 199, 107164.
19. Alesso, C. A., Cipriotti, P. A., Bollero, G. A., & Martin, N. F. (2020). Design of on-farm precision experiments to estimate site-specific crop responses. Agronomy Journal, (December 2020), 1–15.
20. Saikai, Y., Patel, V., & Mitchell, P. D. (2020). Machine learning for optimizing complex site-specific management. Computers and Electronics in Agriculture, 174,
21. Roques, S. E., Kindred, D. R., Berry, P., & Helliwell, J. (2022). Successful approaches for on-farm experimentation. Field Crops Research, 287, 108651.
22. Zhou, X., Heuvelink, G. B. M., Kono, Y., Matsui, T., & Tanaka, T. S. T. (2022). Using linear mixed-effects modeling to evaluate the impact of edaphic factors on spatial variation in winter wheat grain yield in Japanese consolidated paddy fields. European Journal of Agronomy, 133, 126447.
23. Tang, H.;Wang, J.; Xu, C.; Zhou,W.;Wang, J.;Wang, X. Research Progress Analysis on Key Technology of Chemical Fertilizer Reduction and Efficiency Increase. Trans. Chin. Soc. Agric. Mach. 2019, 50, 1–19.
24. Bai, Y.L.; Yang, L.P. Soil Testing and Fertilizer Recommendation in Chinese Agriculture. Soil Fertil. Sci. China 2006, 2, 3–7.
25. Ye, X.; Weng, J. Studies on Fertilization Effect and Recommended Amount for Early Rice Based on“3414” Field Trials. Acta Agric. Univ. Jiangxiensis 2013, 35, 266–273.
26. Singh, M.; Singh, Y.V.; Singh, S.K.; Dey, P.; Jat, L.K.; Ram, R.L. Validation of Soil Test and Yield Target Based Fertilizer Prescription Model for Rice on Inceptisol of Eastern Zone of Uttar Pradesh, India. Int. J. Curr. Microbiol. Appl. Sci. 2017, 6, 406–415.
27. Ahmad, M.; Zahir, Z.A.; Jamil, M.; Nazli, F.; Latif, M.; Akhtar, M.F.-U.-Z. Integrated Use of Plant Growth Promoting Rhizobacteria, Biogas Slurry and Chemical Nitrogen for Sustainable Production of Maize under Salt-Affected Conditions. Pak. J. Bot. 2014, 46, 375–382.
28. Yang, Z.;Wu, X.; Grossnickle, S.C.; Chen, L.; Yu, X.; El-Kassaby, Y.A.; Feng, J. Formula Fertilization Promotes Phoebe Bournei Robust Seedling Cultivation. Forests 2020, 11, 781.
29. Guo, J.; Wang, Y.; Guo, C.; Jin, H.; Yang, Z. Formulas Screening of Special Fertilizer for Spring Maize in County Area of Northern Shanxi Based on GIS and Soil Testing Data. Trans. Chin. Soc. Agric. Eng. 2016, 32, 158–164.
30. Liebig, J.; Von, F.; Playfair, L. Chemistry in Its Application to Agriculture and Physiology. Prov. Med. J. Retrosp. Med. Sci. 1842, 4, 149–152.
31. Mitscherlich, E.A. Des Gesetz Des Minimums Und Das Gesetz Des Abnehmended Bodenertrages. Landwirsch Jahrb 1909, 3, 537–552.
32. Tong, D.; Huang, W.; Ying, R. The Impacts of Grassroots Public Agricultural Technology Extension on Farmers’ Technology Adoption: An Empirical Analysis of Rice Technology Demonstration. China Rural Surv. 2018, 4, 59–73.
33. Harikumaran, M., & Vijayalakshmi, P. (2025). Optimizing fertilizer recommendations in precision agriculture: A novel defuzzification approach with adaptive intelligent optimization. *Knowledge‑Based Systems*, *321*, 113550. <https://doi.org/10.1016/j.knosys.2025.113550>